

Appendix A

GenCore version 5.1.6
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 09:51:41 ; Search time 3878.11 Seconds
 (without alignments)
 9982.023 Million cell updates/sec

Title: US-10-511-270-3
 Perfect score: 1017
 Sequence: 1 cgggatccatgtggggccc.....tgagctgtctcagaattccg 1017

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : EST:
 1: gb_est1:
 2: gb_est2:
 3: gb_htc:
 4: gb_est3:
 5: gb_est4:
 6: gb_est5:
 7: gb_est6:
 8: gb_gss1:
 9: gb_gss2:
 *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	921.2	90.6	1635	3	AK002457	AK002457 Mus muscu
2	908.6	89.3	1596	3	AK010857	AK010857 Mus muscu

ALIGNMENTS

RESULT 1

AK002457	1635 bp mRNA linear HTC 03-APR-2004
LOCUS	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610010D20 product:hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydridopicolinate synthetase containing protein, full insert sequence.
DEFINITION	
ACCESSION	AK002457
VERSION	AK002457.1 GI:12832454
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	(1) Carninci, P. and Hayashizaki, Y.
AUTHORS	
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	(2) Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
AUTHORS	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159

Result 2

LOCUS AK010857 **1596 bp mRNA linear** HTC 03-APR-2004
DEFINITION *Mus musculus* 13 days embryo liver cDNA, RIKEN full-length enriched library, clone:2500002N04 product:hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein, full insert sequence.
ACCESSION AK010857
VERSION AK010857.1 GI:12846588
KEYWORDS HTC; CAP trapper.
SOURCE *Mus musculus* (house mouse)
ORGANISM *Mus musculus*
Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

FEATURES
source Location/Qualifiers
1. .1596
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2500002N04"
/db_xref="taxon:10090"
/clone="2500002N04"
/tissue_type="liver"
/clone_libr="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
28. .1020
/note="unnamed protein product; hypothetical
Aminoacyl-transfer RNA synthetases
class-II/Dihydrodipicolinate synthetase containing protein
(InterPro|IPR002106, InterPro|IPR002220, evidence:
InterPro)
putative"
/codon_start=1
/protein_id="BAB27226.1"
/db_xref="GI:12846589"
/translation="MLGPQIWAWSMROGLSRLSRNVKGKVDIAGIYPPVTPFTATA
EVDYGKLEENLNRLATFPFRGVVQGSTGEFPFLTSLERLEVSVRVRQAIKPDKFLIA
GSGCESTQATVEMTWSMAQVGADVAMVTPCYRGRMSAAIHHHTKVADVSP1PVV
LYSPVANTGLELPVDAVTLSQHPNIIGLKDSGGDVTRIGLIVHKTSKQDPQVLAGSA
GFLLASYAVGAVGGICGLANVLGAQVCQLERLCLTGQWEAAQELQHRLIEPKHCGDPA
LWNTRAEEHGLVWLLWRSLPRPVAGAEPHRGGGTALGFQQOWLALMTSRRRLA"

CDS
ORIGIN
Query Match 89.3%; Score 908.6; DB 3; Length 1596;
Best Local Similarity 94.5%; Pred. No. 4.2e-231;
Matches 952; Conservative 0; Mismatches 54; Indels 1; Gaps 1;
Qy 9 ATGCTGGGCCCAATCTGGCCTCCATGAGGCAGGGCTGAGCACAGGGCTTGCTAGG 68
Db 28 ATGCTGGGCCCAATTGGCCTCCATGAGGCAGGGCTGAGCACAGGGCTTGCTAGG 87
Qy 69 AACGTGAAGGGAAAGAGATAGACATTCGGCCATCTACCCACCCGTGACCACCCATTC 128
Db 88 AATGTGAAGGGCATGAAGGTAGACATTCGGCCATCTACCCACCCGTGACCACCCATTC 147
Qy 129 ACCGCCACCGCAGAAGTAGACTATGGGAAACTGGAAAGAGAACCTGAACAAACTGGCC 188
Db 148 ACCGCCACCGCAGAGGTAGACTATGGGAAACTGGAAAGAGAACCTGAACAGACTGGCCACC 207
Qy 189 TTCCCCTTCGAGGCTTCGTGGTCAGGGCTACTGGAGAGTTCCATTCTGACCAGC 248
Db 208 TTCCCCTTCGAGGCTTCGTGGTCAGGGTTGACTGGAGAGTTCCGTGACCAGC 267
Qy 249 CTTGAGCGCCTAGAGGTGGTGAGCGAGTCGGCCAGGCCATACCCAGGACAAGCTCC 308
Db 268 CTCGAGCGCTGGAGGTGGTGAGCGCGTGCAGGCCATACCCAGGACAAGCTCC 327
Qy 309 ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT 368
Db 328 ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT 387
Qy 369 CAGGTGGGTGCTGATGCCCATGGTGGTGACCCCTTGTACTATCGCGCCGATGAAC 428
Db 388 CAGGTGGGTGCTGATGCCCATGGTGGTGACCCCTTGTACTATCGCGCCGATGAGC 447

Qy	429	AGCGCTGCCCTCATTCAACCACTACACCAAGGTTGCTGATCTTCTCCAATCCGGTGGTG	488
Db	448	AGCGCTGCCCTCATTCAACCACTACACCAAGGTTGCTGACGTTCTCCAATCCCTGTTG	507
Qy	489	CTGTACAGTGTCCCAGGCAACACGGGCTCTAGAGCTGCCTGTGGATGCCGTGTCACATTG	548
Db	508	TTGTACAGTGTCCCAGGCAATACGGGCTAGAGCTACCTGTGGATGCCGTGGTTACATTG	567
Qy	549	TCTCAGCACCCAAATATCATTGGCTTGAGGACAGTGGTGGAGATGTGACCAGGACTGGG	608
Db	568	TCTCAGCACCCAAATATCATCGGCTTGAGGACAGTGGTGGAGATGTGACCAGGATGGG	627
Qy	609	CTGATTGTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTCAAGTTGGCTTC	668
Db	628	CTGATAGTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTCAAGCTGGCTTC	687
Qy	669	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGCATATGTGGCTGGCAATGTCTTG	728
Db	688	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGCATATGTGGCTGGCAATGTCTTG	747
Qy	729	GGGGCCCAGGTGTGCCAGCTGGAGAGACTCTGCCCTCACAGGGCAGGGGGAAAGCTGCCAG	788
Db	748	GGGGCCCAGGTGTGCCAGCTGGAGAGACTCTGCCCTCACAGGGCAGTGGAAAGCTGCCAG	807
Qy	789	AGACTGCAGCACCGCTCATCGAGCCC-AACACTGCGGTGACCCGGCGTTTGGAAATACC	847
Db	808	GAACTACAGCACCGCTCATCGAGCCAAACACTGCGGTGACCCGGCGTTTGGAAATACC	867
Qy	848	AGGGCTGAAGAAAACCATGGACTGGTTGGCTACTATGGAGGTCCCTGCCGTGCCCTT	907
Db	868	AGGGCTGAAGAAAACCATGGACTGGTTGGCTACTATGGAGGTCCCTGCCGTGCCCTT	927
Qy	908	GCAGGAGTTGAGCCCCCTCAGAGGAAGAGGCGCTTCGCTTGGATTTCAGCAACAATGGCTG	967
Db	928	GCAGGAGCTGAGCCCCACAGAGGGAGGCACCTGCGCTTGGATTTCAGCAACAATGGCTG	987
Qy	968	GCTTTAATGACAAGCGGGGACACCTGGCTGAGCTGTCTCAGAATT	1014
Db	988	GCTTTAATGACAAGCAGGAGACGCCCTGGCTGAGCTATCTGGACTT	1034